

## The God / Ego Hypothesis

Consciousness has always been one of the great problems of both science and philosophy. How is it that I see; that I smell; that I exist at all? What is it to be an 'I' and are you one too? What is the ego? We have thought about it a lot and came up with a lot of theories we divided into two groups: there is dualism which says the secret of consciousness lies in an immaterial soul and we have materialism where everything is mechanical. The origin of this consciousness and the self is something I'm going to try to speculate on with shameless arrogance.

I think the whole idea of an immaterial soul (what does that really even mean?) is as schizophrenic as melting walls at the height of an acid trip so we'll go with materialism and mechanics for now. It should be rather obvious that the procedure I'm about to describe would also function with an evolutionary process (probably a lot easier given enough time) but let's say we are going to play God and 'create life'. Because we petty humans lack the skill of properly wielding organic material, we are going to create this lifeforms in a digital world where it can freely move around without all these problems with moving muscles and so forth. Our lifeform lives in a four dimensional world much like our own (three spatial and 1 time dimension) and has a body with which he can move around in this world the same way you and I do, let's say it's also a humanoid.

But we aren't as much concerned with his physical body, we are out for his consciousness, as Turing put it, the way we see it reacting. To create this consciousness we will implement an artificial intelligence that does 3 things: receive data from various senses in the body; filter and analyze the data for patterns and react in the external world. It will also be initialized with a couple of 'hardcoded' constants\* for the determining the final value of the data it is analyzing, in an evolutionary process they would start random and be optimized vis a vis survival. These constants thus act as basic security mechanisms for operating the body, think of something like physical pain. We assume that these constants are relatively functional for keeping the lifeform alive in the environment as long as nothing exciting happens. As it is analyzing the data it's receiving from it's senses it strives to generalize as much patterns as possible. If there is a pattern that appears a lot, this is stored in memory. By doing this the lifeform builds a virtual world encoded inside it's neural network, which isn't as filled with valuable information about the external reality as the real thing but does have the bonus of being persistent. It can now find out the appropriate reaction (optimized for it's values around that moment) to certain patterns that appears more often than normal and store this information.

By doing this the lifeform can learn causality by itself. It does that by encoding causality in sets of multiple patterns. If it sees patterns A and B together often, then it might group these patterns into a set of patterns. Let say it picks up a stone and drops it, it goes downwards to the surface. Those are two patterns that appear every time you try it. But let say it didn't try to drop a stone next time, but a log of wood. Again it goes downwards to the surface, but this wasn't a stone. Ok, so now it can conclude that both 'log of wood' and 'stone' (or object X and Y whatever you prefer) are connected to the 'go downwards to surface' pattern. This means the set containing them exists and 'go downwards to surface' is a property linked to this set. It should be rather obvious that if you gather a lot of data with different interactions with the external world that a huge network of interconnected patterns and sets of patterns would be the result. This is actually how our lifeform will become aware of the generalization 'object', it is a very large set with few properties.

Whilst chilling in this external world the lifeform sometimes comes across another lifeform. It has distinguished them from other things because other lifeforms appear as an enormous amount of patterns changing rapidly within a small set. Their 'body' (pattern) is itself a surface of patterns that appears to

be rather stable in the three dimensions of space, even though they are changing constantly it is always within a very limited or specialized set. But over time it is connected with an enormous amount of other patterns. It seems to have a lot of causal links to other patterns that are itself a lot less active (which we call inanimate 'objects'). We will crown the surface containing all these patterns and the way they interact with other patterns in the 4<sup>th</sup> dimension of time 'person'.

So far so good but I have the feeling we missed something. How does our lifeform receive information from the outside world? It is data received with sensors connected directly to the outside world, which we will call 'nerves'. These nerves aren't equally spread over it's body but they can be grouped into a set of senses. These senses are then filtered through some rather constant mechanisms and send to the neural network for analyzing. This means that our lifeform could perceive something like visual stimulation, which would in these terms mean that a small concentrated group of nerves would receive data from being hit particles / waves scattering off different objects in the external world. For instance it could measure the intensity of small sets of nerves to measure the direction of the light, the intensity of larger sets for the actual intensity of the light. The visual image of another object or person in this external world would be a pattern or set of patterns found by analyzing this nerve data.

Ok so our lifeform thinks it's a smart motherfucker and groups other lifeforms into 'persons'. But it doesn't know that it's a lifeform itself yet, that 'it' exists. It's just analyzing data with some lifeless neural network and acting upon in the external world, doing just what it remembered working best last time a pattern like this came around. But if it groups other lifeforms into persons because it detects other lifeforms with it's senses, wouldn't it then be logical for our lifeform to group it's own body into a 'person'? Since it's own body would probably have the most effect on it's senses during the time that it exists, it has the acquired the most data of it and thus is aware of a lot of the generalizations in the patterns.

It is now not only what we perceive as life when thinking like Turing, analyzing and acting already pretty much covered that, but it is now also to our most basic standards 'a person'. While it is still a purely mechanical body with a neural network trained to analyze and generalize patterns on which it's acts, his virtual world now also contains a person describing it's own body. It's still not aware that this is actually 'itself' but it resembles other persons in that it's a rather constant surface of patterns in the third dimension, but changes quite a lot over a small period of time (within a small set of patterns). Also there is a huge list of causal links to other patterns. And though very small and weak, the ego is born.

Since it is trying to continue it's existence (which is the goal of the constant values), it learns a lot of causal links between changes in it's body. Especially the configuration of different smaller patterns within the body in the third dimension (i.e limbs moving around) and it's own reaction. It doesn't necessarily know that these limbs are it's own limbs, because it doesn't know it exists that much yet. What it does know, is that for some reason it can control the patterns of these limbs (moving them around). Lots of experimentation and acquiring data will have the effect that our lifeform learns a lot about it's own body and how changing different patterns make his neural network react.

Now thus far it all sounds pretty solid if I may say so myself but this is where it really gets speculation. It should also be noted that the computing power of our lifeform isn't endless. We were rather cheap and choose a shared hosted server for this little buddy so he will have to do with a limited amount of resources. Especially time is precious, for it has no causal patterns with which it can change time for the better, it just keeps ticking away. This means that there are certain operations that take a longer time than others. Because our lifeform is trying to be an efficient guy he has a some very high level generalized causality rules, called 'logic'. Also in the meantime he has been chilling a lot with other lifeforms and they have developed a way of communicating by making sounds with their body. Maybe it has also discovered that if it really wants to know something as sure as possible, it's handy to call a very high level logic check by encoding an internal state into language. This is an operation that is very expensive and should only be used if heuristics have indicated that it's worth doing that. This process is what the lifeform communicates to other lifeforms encoded by the sound 'thoughts'.

This will also help building the 'ego' person, because optimization means that idle time should also be used properly. When it has no immediate threats from the outside world it can generate permutations of information and check the results with heuristics / logic. Because it marked the ego person as very influential on his virtual world, this is a logical subject for idle time analyzing. It should also be noted that if there are a lot of these lifeforms around and they have a way of communicating pretty abstract information that this process will only go faster and faster. The more it communicates with other lifeforms the more they will appear similar to data it has acquired about its own body. This will have an strengthening effect on the ego, the hallucination of itself.

To go even further into my arrogance and speculate on the thing thou shall not speculate about! Yes I'm speaking of the one and only, God. Our lifeform also has a God because it has grown quite fond of this person configuration. It turns out to be an incredibly helpful tool in the quest of discovering what the fuck is going on. Now life wouldn't be fun if it wasn't zero-sum, so that's also the way our lifeform sees it. Everything that it perceives is a set. Sets containing smaller sets, containing EVEN SMALLER SETS (and yes indeed, even smaller sets), but eventually also the 'universe' is just the powerset of everything it knows. This also means that there is a set of patterns that appear rather often or have a huge impact on the virtual world of our lifeform but don't really seem to have any causal links. It can be spotted individually quite clearly but there is no data whatsoever about what is causing it, so that is the common factor. There is no causal link at all, this will define our set. This is a set we can proudly call God, and it wouldn't be God if we wouldn't qualify this set as a person. God is a weird person, just like the ego is a very weird person as they do not seem to behave quite like other persons when it comes to what is in the set, but how the set as a whole behaves is pretty remarkably similar. The body or plane of patterns (or large set containing lots of other sets, for yet another term) itself is rather constant: God has no causality and the self gives a whole range of unique physical feedback. This means that it's aware of itself and of a God, qualities usually ascribed to the highest of conscious beings, yet all this behavior is derived from a purely mechanical composition.

The weird thing is that it sounds pretty simple to me and I'm not sure that I'm right, but I cannot find any logical flaws in this while assuming the few things I had to assume. I'm not an expert in artificial intelligence (also not a complete nitwit), but most of what I'm describing should, to my knowledge, actually be possible with neural networks and evolutionary programming. So please prove me wrong because I'm not willing to go all-in on this, but I also cannot find any flaws myself.

*\*It should be noted that 'constant' doesn't necessarily have to mean CONSTANT, but it should at least be very fucking hard to change. In the physical world these constants could be encoded into genetics or learned in a prenatal or young state.*

Tim Jansen 20-03-2012